Die 'Dresden Image Database' für die Entwicklung und Validierung von Methoden der digitalen Bildforensik

Thomas Gloe
Technische Universität Dresden
Fakultät Informatik
01062 Dresden

Thomas.Gloe@tu-dresden.de

Abstract: Dieser Aufsatz stellt eine neue Bilddatenbank zur Entwicklung und Validierung von Methoden der digitalen Bildforensik mit Fokus auf Digitalkameraaufnahmen vor. Die Datenbank wurde unter kontrollierten Bedingungen unter Verwendung von insgesamt 73 verschiedenen Digitalkameras erstellt. Sie beinhaltet mehr als 15.000 unterschiedliche Bilder von Innen- und Außenszenen. Die verwendeten Digitalkameras sind jeweils einem von 25 verschiedenen Kameramodellen zuzuordnen, wobei zur Erstellung der Aufnahmen bis zu fünf Geräte pro Modell verwendet wurden. Die Bilddatenbank ermöglicht somit sowohl eine Untersuchung von modellabhängigen Charakteristiken anhand mehrerer Geräte eines Modells als auch die Untersuchung von spezifischen Charakteristiken eines einzelnen Geräts. Zusätzlich wurden spezielle Aufnahmen für die Untersuchung des geräteabhängigen Sensorrauschens und für die Betrachtung von modellabhängigen JPEG Kompressionsartefakten erstellt. Die Verwendung einer Referenz-Datenbank für die Entwicklung und Validierung von Methoden der digitalen Bildforensik macht zukünftige Forschungsergebnisse nicht nur besser vergleichbar und reproduzierbar, sie vermeidet auch Urheberrechtsverletzungen und Eingriffe in die Privatsphäre bei der Verwendung von öffentlich zugänglichen Bildern aus Foto-Communities im Internet, wie beispielsweise Flickr[1]. Die 'Dresden Image Database' ist für wissenschaftliche Zwecke öffentlich frei zugänglich und soll somit eine wichtige Forschungsgrundlage für Wissenschaftler und Praktiker auf dem Gebiet der digitalen Bildforensik werden.

1 Motivation

Die heute verfügbaren digitalen Bildeingabegeräte und die dazugehörigen Werkzeuge ermöglichen eine schnelle und hochwertige Erstellung von digitalen Bildern. Mit dem technischen Fortschritt stieg auch die Nutzerakzeptanz, so dass digitale Bilder einen vielfältigen Einsatz in allen Lebensbereichen finden, sei es in Form von privaten Schnappschüssen, als Ergebnis professioneller Fotoarbeiten oder bei der Überführung von analogen Dokumenten in die digitale Welt. Dabei macht die Verwendung von digitalen Bildern als Beleg in den Medien, in wissenschaftlichen Publikationen, vor Gericht oder bei Schadensmeldungen gegenüber Versicherungen, die Entwicklung geeigneter Methoden zur Erken-

[1] www.flickr.com

nung von Bildmanipulationen erforderlich. Zusätzlich können digitale Bilddaten wichtige Informationen über das verwendete Bildeingabegerät liefern und somit einen Hinweis auf verdächtige Personen in forensischen Untersuchungen geben. Die beiden genannten Aspekte werden im Forschungsgebiet der digitalen Bildforensik genauer untersucht.

Derzeit sind bereits eine Vielzahl von verschiedenen wissenschaftlichen Publikationen und unterschiedlichen Ansätzen im Gebiet der digitalen Bildforensik bekannt [PF04, KSM04, JF05, LFG05, JF06, NCLS06, u.a.]. Die veröffentlichten Experimente und Untersuchungsergebnisse basieren dabei meist auf Datensätzen einzelner Forschergruppen und sind für freie Forschungszwecke in der Regel nicht öffentlich zugänglich[2]. Das Fehlen einer allgemeinen Referenzbilddatenbank erschwert damit sowohl die Vergleichbarkeit als auch die Reproduzierbarkeit von Forschungsergebnissen. Die daraus folgende Notwendigkeit einer allgemein zugänglichen Bilddatenbank wurde in der Literatur bereits von verschiedenen Autoren angemerkt [SM08, NCLS06]. Hinzu kommt, dass die Erstellung von Bilddatenbanken durch jede einzelne Forschergruppe aus forschungsökonomischen Gesichtspunkten wenig effizient erscheint. Um diese Lücke zu schließen, möchten wir in diesem Aufsatz die 'Dresden Image Database' vorstellen.

Der Aufbau der Bilddatenbank wurde unter größter Sorgfalt und unter Beachtung der Anforderungen von Wissenschaftlern und Praktikern auf dem Gebiet der digitalen Bildforensik entwickelt. Ziel war es eine umfassende und erweiterbare Datenbank von Aufnahmen realer Szenen zu erstellen. Um die derzeit bekannten Einflussfaktoren gezielt untersuchen zu können, wurden alle Bilder unter kontrollierten Bedingungen aufgenommen. Insbesondere wurde bei der Auswahl von Szenen auf Urheberrechte und Rechte der Privatsphäre geachtet. So werden beispielsweise bei Untersuchungen eines digitalen Bildes zur Erkennung der verwendeten Digitalkamera zusätzliche Bilder von anderen Geräten als negative Testbeispiele benötigt. Kommen in diesem Zusammenhang Bilder aus öffentlich zugänglichen Fotocommunities zum Einsatz, ist allerdings die Frage ungeklärt, wie mit einem Treffer in der Menge der negativen Testbeispiele umgegangen werden soll und inwieweit der Autor des betreffenden Bildes unter Verdacht gerät. Diese Probleme treffen nicht auf unsere Bilddatenbank zu.

Ausgehend von einer Beschreibung der Entwurfsentscheidungen und der Prozedur zur Erstellung der Aufnahmen für die Bilddatenbank in Abschnitt 2, gibt Abschnitt 3 einen Überblick über den Inhalt der Bilddatenbank einschließlich ausgewählter Statistiken. In Abschnitt 4 werden zusätzlich bereitgestellte Bilder für die Untersuchung des geräteabhängigen Sensorrauschens und die Untersuchung von modellabhängigen JPEG Kompressionsartefakten vorgestellt. Ausgewählte Analyseergebnisse am Beispiel der Identifikation von Kameramodellen in Abschnitt 5 belegen die Nutzbarkeit der Bilddatenbank. Abschnitt 6 gibt eine Zusammenfassung dieses Aufsatzes.

[2] Eine Ausnahme bildet die auf Image Splicing Operationen spezialisierte Bilddatenbank von Hsu und Chang [HC06]. Sie ist allerdings auf sehr einfache Manipulationen beschränkt und beinhaltet nur einen kleinen Bilddatensatz.

2 Erstellung der Bilddatenbank

Ziel der Bilddatenbank ist es sowohl modell- als auch geräteabhängige Charakteristiken unter idealen Bedingungen anhand von unverfälschten Aufnahmen studieren zu können. Dies erscheint besonders gut möglich, wenn Aufnahmen natürlicher Szenen mit sehr ähnlichem Bildinhalt für alle Digitalkameras in der Bilddatenbank vorliegen. Die ausgewählten natürlichen Szenen wurden deshalb mit jeder Digitalkamera aufgenommen. Zusätzlich beeinflussen Kameraeinstellungen das Ergebnis einer Aufnahme. Bei der Erstellung der Bilddatenbank wurden deshalb einzelne Kameraeinstellungen gezielt variiert.

Insgesamt wurden 25 verschiedene Kameramodelle bekannter Hersteller für die Erstellung der Bilddatenbank ausgewählt. Dabei standen pro *Kameramodell* bis zu fünf verschiedene *Geräte* zur Verfügung. Die verwendeten Kameramodelle schließen das gesamte Preis- und Leistungsspektrum von sehr einfachen Geräten mit fester Brennweite bis zu semi-professionellen Spiegelreflex-Digitalkameras ein. Dies ermöglicht die Entwicklung und Überprüfung von Algorithmen anhand der gesamten Bandbreite von typischerweise eingesetzten Digitalkameras. Für die praktische Realisierung der Bilddatenbank war eine Trennung der Vielzahl von Geräten in zwei Kameramengen notwendig. Es wurde allerdings sichergestellt, dass die Geräte des gleichen Modells der gleichen Kameramenge zugeordnet sind. Somit liegen von allen Geräten des gleichen Modells auch Aufnahmen der gleichen Szene vor.

Die Auswahl der Szenen für die Erstellung der Bilddatenbank erfolgte unter Einschluss von Nah- und Fernaufnahmen in der Umgebung von Dresden. In Abbildung 1 sind typische Beispiele für die erstellten Aufnahmen dargestellt. Für das Fotografieren der verschiedenen Szenen wurden unterschiedliche Aufnahmestandorte eingenommen und jeweils ein festgelegter Versuchsaufbau installiert. Abbildung 2 stellt diesen vereinfacht dar. Generell wurden an jedem *Standort* zwei unterschiedliche *Motive* unter Verwendung von zwei fest installierten Stativen fotografiert. Von jedem Motiv wurden drei oder mehr Aufnahmen durch Veränderung der Brennweite erstellt. Dabei wurde bei jeder Kamera sowohl die minimale als auch die maximale Brennweite verwendet und zusätzlich die Szenen mit einer zwischen Kameramodellen möglichst vergleichbaren Brennweite fotografiert. Um die Anzahl der erstellten Aufnahmen weiter zu erhöhen, wurde bei der Erstellung der Aufnahmen für die zweite Kameramenge zusätzlich eine vierte Brennweite für jede Kamera verwendet. Beim Fotografieren der Szenen wurde entweder der Automatik- oder, falls verfügbar, der Programmmodus aktiviert. Außerdem wurde die Blitzautomatik der Kameras verwendet und beim Auslösen des Blitzes eine zusätzliche Aufnahme mit deaktiviertem Blitz erstellt.

Die meisten gängigen Digitalkameras erlauben lediglich das Speichern von komprimierten Bilddaten im JPEG Format. Um den Einfluss von Bildverarbeitungsartefakten durch JPEG Kompression zu minimieren, wurde in allen Kameras die höchste Qualitätsstufe für die Erstellung der Aufnahmen verwendet. Soweit es einzelne Kameras ermöglichen, wurden die Aufnahmen zusätzlich in einem verlustfrei kodierten Format gespeichert.

Da die Erstellung aller Aufnahmen mit allen Kameras einer Kameramenge an einem Standort meist innerhalb einer Stunde umgesetzt werden konnte, war es möglich, die Be-

Abbildung 1: Beispiele für unterschiedliche Innen- und Außenaufnahmen von natürlichen und urbanen Szenen in der Bilddatenbank. Alle Kameras von Kameramenge A (vgl. Tab. 1) liegen im linken unteren Bild auf dem Schreibtisch.

leuchtungsbedingungen pro Szene vergleichsweise konstant zu halten. In einigen wenigen Fällen führten allerdings verschiedene Umwelteinflüsse, wie schnelle Wetteränderungen, zu Variationen in der Beleuchtungssituation der einzelnen Aufnahmen. Für ausgewählte Standorte wurden deshalb zusätzliche Kontrollaufnahmen am Anfang und am Ende der gesamten Aufnahmeprozedur mit einer ausgewählten Kamera erstellt. Dies ermöglicht eine Berücksichtigung von Veränderungen in der Beleuchtungssituation während der gesamten Aufnahmezeit.

Die Zuordnung der erstellten Aufnahmen zu einem Motiv und zu einem Standort erfolgt innerhalb der Bilddatenbank anhand von verschiedenen Zusatzinformationen. Dies vereinfacht sowohl die Auswahl bestimmter Aufnahmen mit bestimmten Eigenschaften als auch den Vergleich von Aufnahmen hinsichtlich des Einflusses von Kameraeinstellungen, modell- und geräteabhängigen Charakteristiken.

3 Details der Bilddatenbank

Tabelle 1 gibt einen Überblick über die bei der Erstellung der Bilddatenbank verwendeten Kameramodelle. Zusätzlich sind Informationen zur Anzahl der verwendeten Einzelgeräte, zu den Kameraspezifikationen und zur Anzahl der erstellten Bilder mit und ohne aktiviertem Blitz angegeben. In Kameramenge A (Set A) wurden insgesamt 47 unterschiedliche

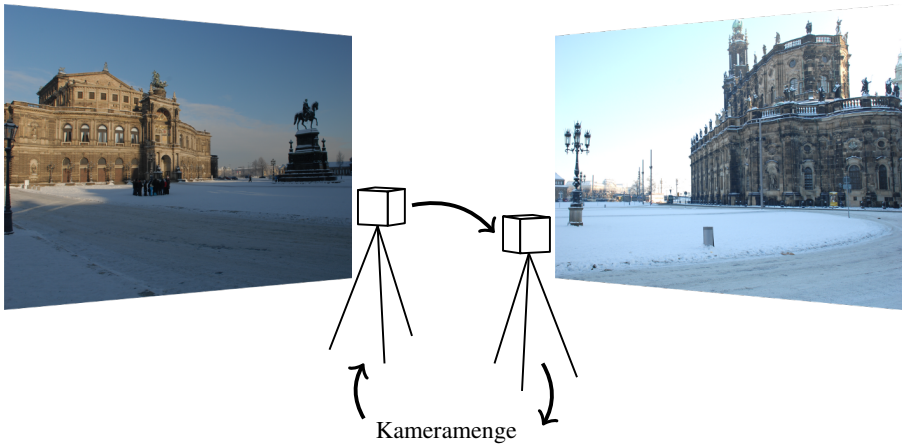

Abbildung 2: An jedem Aufnahmestandort wurden 2 Stative aufgestellt und anschließend mit jeder Kamera einer Menge beide Szenen mit unterschiedlichen Kameraeinstellungen fotografiert.

Szenen und in Kameramenge B (Set B) 30 unterschiedliche Szenen fotografiert. Die Anzahl der bereitgestellten Bilder in der Datenbank ermöglicht generell auch eine Untersuchung von Algorithmen unter festgelegten Kameraeinstellungen.

Außerdem wurde darauf geachtet, dass die Variabilität von Abbildungsfehlern in natürlichen Szenen möglichst genau untersucht werden kann. Hierfür wurden zwei identische Wechselobjektive in Kombination mit den vier zur Verfügung stehenden digitalen Spiegelreflexkameras vom Typ Nikon D70/D70s verwendet. Jedes der beiden Wechselobjektive war dabei zwei der vier Geräte zugeordnet und jede Szene wurde mit jedem Gerät und einem der beiden zugeordneten Objektive erstellt. Außerdem wurden beim Kameramodell Nikon D200 zwei unterschiedliche Wechselobjektive verwendet, die für jede Szene zwischen den beiden verwendeten Geräten ausgetauscht wurden. Im Vergleich zu den anderen Kameramodellen entstanden somit bei der Nikon D200 die doppelte Anzahl von Bildern pro Szene und pro Gerät.

Eine Möglichkeit, die Qualität der Bilddatenbank zu untersuchen, ist den Anteil von gesättigten Bildbereichen in den erstellten Aufnahmen zu bestimmen. Dabei gelten Bereiche als gesättigt, wenn sie entweder nur sehr geringe Werte im Bereich von 0 oder nur sehr hohe Werte im Bereich von 255 aufweisen[3]. Die Sättigung entsteht meist durch Unter- oder Überbelichtung von Teilen einer Szene. Zwar lässt sich die Sättigung in vielen kontraststarken Szenen nicht ausschließen, aber eine Auswertung von geräte- oder modellabhängigen Charakteristiken ist in diesen Bereichen nur eingeschränkt möglich. Abbildung 3 zeigt die Zahl von Sättigungsbereichen in Kameramenge A als Histogramm über dem Verhältnis von Anzahl von gesättigten Bildpunkten eines Bildes zur Gesamtanzahl der vorhanden Bildpunkte. Für die Berechnung des Histogramms wurde von allen Bildern zunächst die

[3]Die Zahl 255 bezieht sich dabei auf den maximalen Wert bei Bildern mit 8-bit Farbtiefe.

Tabelle 1: Überblick über die verwendeten Digitalkameramodelle und deren Eigenschaften (Anzahl von Geräten pro Modell, grundlegende Kameraspezifikation und Anzahl bereitgestellter Aufnahmen).

Kamera Modell	Anzahl Geräte	max. Auflösung [Bildpunkte]	Sensorgröße [inch]	Brennweite [mm]	Anzahl Bilder Set A (Blitz aus/an)	Anzahl Bilder Set B (Blitz aus/an)
AgfaPhoto DC-504	1	4032×3024	-	7.1	84 (70/14)	80 (70/10)
AgfaPhoto DC-733s	1	3072×2304	-	6.2−18.6	151 (130/21)	166 (163/3)
AgfaPhoto DC-830i	1	3264×2448	-	6.2−18.6	183 (132/51)	217 (183/34)
AgfaPhoto Sensor 505-X	1	2592×1944	-	7.5	87 (76/11)	104 (95/9)
AgfaPhoto Sensor 530s	1	4032×3024	-	6.1−18.3	199 (149/50)	198 (160/38)
Canon Ixus 55	1	2592×1944	1/2.5”	5.8−17.4	186 (143/43)	-
Canon Ixus 70	3	3072×2304	1/2.5”	5.8−17.4	542 (428/114)	-
Casio EX-Z150	5	3264×2448	1/2.5”	4.65−18.6	925 (748/177)	-
Fujifilm FinePix J50	3	3264×2448	1/2.5”	6.2−31.0	-	495 (433/62)
Kodak M1063	5	3664×2748	1/2.33”	5.7−17.1	999 (692/307)	1161 (920/241)
Nikon Coolpix S710	5	4352×3264	1/1.72”	6.0−21.6	916 (743/173)	-
Nikon D70/D70s	2/2	3008×2000	23.7×15.6 mm	18−200	715 (637/78)	-
Nikon D200 Lens A/B	2	3872×2592	23.6×15.8 mm	18 − 135/17 − 55	712 (633/79)	-
Olympus μ1050SW	5	3648×2736	1/2.33”	6.7−20.1	1027 (686/341)	-
Panasonsic DMC-FZ50	3	3648×2736	1/1.8”	7.4−88.8	-	561 (493/68)
Pentax Optio A40	4	4000×3000	1/1.7”	7.9−23.7	-	629 (540/89)
Pentax Optio W60	1	3648×2736	1/2.3”	5.0−25.0	-	146 (129/17)
Praktica DCZ 5.9	5	2560×1920	1/2.5”	5.4−16.2	973 (706/267)	-
Ricoh Capilo GX100	5	3648×2736	1/1.75”	5.1−15.3	-	715 (632/83)
Rollei RCP-7325XS	3	3072×2304	1/2.5”	5.8−17.4	581 (433/148)	-
Samsung L74wide	3	3072×2304	1/2.5”	4.7−16.7	562 (425/137)	-
Samsung NV15	3	3648×2736	1/1.8”	7.3−21.9	566 (456/110)	-
Sony DSC-H50	2	3456×2592	1/2.3”	5.2−78.0	-	308 (284/24)
Sony DSC-T77	4	3648×2736	1/2.3”	6.18−24.7	-	565 (506/59)
Sony DSC-W170	2	3648×2736	1/2.3”	5.0−25.0	-	272 (246/26)
$\sum$	73				9408 (7287/2121)	5617 (4854/763)

Helligkeitsinformation bestimmt. Anschließend wurden Bildpunkte als gesättigt gezählt, wenn ihre Werte entweder kleiner oder gleich 5 oder größer oder gleich 250 sind. Die Schwellwerte entsprechen dabei dem typischen Rauschlevel einer Digitalkamera. In der Bilddatenbank liegen bei 97% aller Bilder von Kameramenge A weniger als 25% der Bildpunkte im Sättigungsbereich und bei 89% der Bilder weniger als 10% der Bildpunkte im Sättigungsbereich. Genauer ausgedrückt sind bei 95% der Bilder weniger als 10% der Bildpunkte in den hellen Bereichen (≥ 250) gesättigt und bei 94.5% der Bilder weniger als 10% der Bildpunkte in den dunklen Bereichen (≤ 5) gesättigt.

4 Zusätzliche Spezialaufnahmen zur Untersuchung von ausgewählten Charakteristiken

Neben den Aufnahmen von natürlichen Szenen wurden spezielle Fotografien für die Untersuchung des Sensorrauschens und die Untersuchung von JPEG Kompressionsartefakten erstellt.

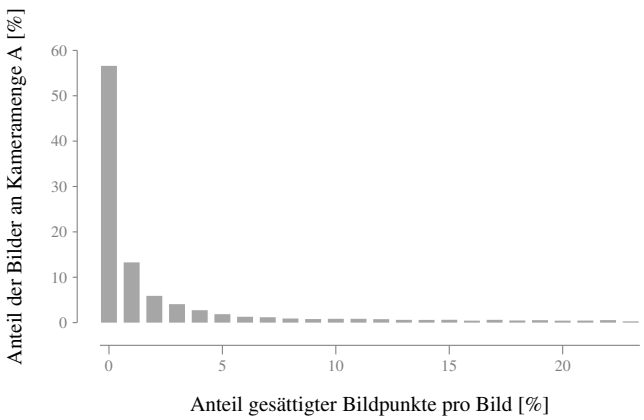

Abbildung 3: Histogramm der Sättigungsbereiche infolge von Unter- oder Überbelichtung in Kameramenge A. Die Darstellung zeigt 97% aller Bilder. Die restlichen Bilder enthalten Sättigungsbereiche mit einem Anteil von mehr als 24% der Bildpunkte pro Aufnahme.

4.1 Sensorrauschen

Generell beinhaltet das Sensorrauschen sowohl stabile als auch sehr variable Rauschanteile. Bei der Nutzung des Sensorrauschens zur Identifikation einer Kamera oder zur Erkennung von Bildmanipulationen werden die stabilen Rauschanteile eines Geräts im Bild ausgewertet [LFG05, LFG06a]. Für eine genaue Untersuchung der einzelnen Bestandteile des Sensorrauschens der Kameras in der Bilddatenbank wurden deshalb 50 Dunkel- und 50 Hellaufnahmen für jedes Gerät erstellt. Generell ermöglichen Dunkel- und Hellaufnahmen ein genaue Schätzung der beiden stabilen Bestandteile des Sensorrauschens: Dunkelrauschen (Fixed Pattern Noise) und Photoresponse Non-Uniformity. Die Dunkelaufnahmen wurden durch Abdecken der Linse erstellt und messen somit das Dunkelrauschen, welches allein durch die elektronischen Bauelemente ohne den Einfall von Licht entsteht.

Ziel der Hellaufnahmen ist es, die Photo-Response Non Uniformity einer Kamera zu quantifizieren. Diese entsteht durch Unterschiede in der Helligkeitsempfindlichkeit einzelner Sensorelemente und kann unter Verwendung einer homogen beleuchteten Glasscheibe besonders gut gemessen werden. Zur Erstellung der Hellaufnahmen wurde deshalb der Durchlichtaufsatz eines Epson Perfection 1240U Flachbettscanners verwendet. Die erstellten Aufnahmen ermöglichen auch weiterführende Studien zu Möglichkeiten der Fälschbarkeit des Sensorrauschens (zum vgl. [LFG06b, GKWB07]).

4.2 JPEG Kompressionsartefakte

Ein weiterer Bilddatensatz wurde speziell für die Untersuchung von JPEG Kompressionsartefakten entwickelt. Je nach Digitalkamerahersteller kommen entweder Kompressi-

Abbildung 4: Die beiden Spezialszenen zur genauen Untersuchung der kameramodellabhängigen JPEG Kompressionsalgorithmen.

onsalgorithmen zum Einsatz, die unabhängig vom Szeneninhalt arbeiten (beispielsweise durch Auswahl einer bestimmten Kompressionsqualität), oder es kommen Kompressionsalgorithmen zum Einsatz, die die JPEG Kompression entsprechend der Szenenkomplexität anpassen (beispielsweise um eine bestimmte Dateigröße sicherzustellen). Um diesen Sachverhalt genauer zu untersuchen und um eine relevante Datenbank von Quantisierungstabellen pro Kameramodell zur Verfügung zu stellen, wurden zwei unterschiedliche Szenen mit unterschiedlicher Szenenkomplexität entwickelt. Abbildung 4 zeigt zwei Aufnahmen der beiden entwickelten JPEG Szenen. Die erste Szene enthält viele homogene Bildbereiche und somit nur wenige Wechsel in den Intensitätswerten. Im Gegensatz dazu ist die zweite Szene deutlich komplexer und enthält viele Farb- und Kontrastwechsel. Da die Implementierung der JPEG Kompression für verschiedene Geräte eines Kameramodells mit gleicher Firmeware gleich ist, wurden die beiden Szenen mit je einem Gerät eines Kameramodells aufgenommen. Dabei wurde über alle möglichen Kombinationen von Bildgröße, Kompressionsqualität und mit/ohne Blitz iteriert.

5 Erkennung des verwendeten Kameramodells

Für ein erstes Experiment wurde die Leistungsfähigkeit der Methode von Kharrazi, Sencar und Memon zur Bestimmung des verwendeten Kameramodells auf Basis der Bilddatenbank untersucht [KSM04]. Generell basiert diese Methode auf einer Auswertung verschiedener grobgranularer Merkmale, die durch modellabhängige Charakteristiken der Farbwiedergabe, der Abbildungsqualität und der Rauscheigenschaften bestimmt sind. Zur Zuordnung eines Bildes zum vermutlich verwendeten Kameramodell muss zunächst das gewählte Klassifikationsverfahren (bspw. eine Support Vector Machine [CL01]) mit den berechneten Merkmalen einer Lernbildmenge trainiert werden. Anschließend kann der angelernte Klassifikator für die Zuordnung von Bildern zu den vermutlich genutzten Kameramodellen verwendet werden. Im Gegensatz zu den bisher bekannten Untersuchungsergeb-

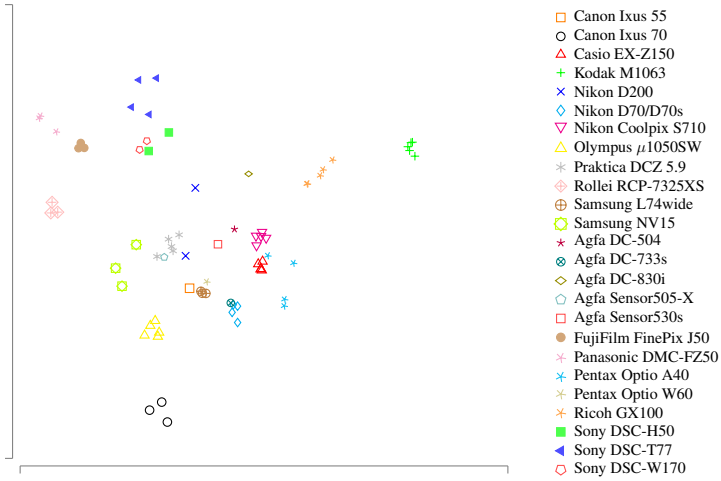

Abbildung 5: Veranschaulichung der Ähnlichkeit zwischen den berechneten Merkmalen von Kameras eines Modells und der deutlichen Trennbarkeit zwischen unterschiedlichen Kameramodellen. Für die Darstellung wurde die Mittelwerte der 10 wichtigsten Merkmale aller Bilder eines Geräts mit multidimensionaler Skalierung in einen zweidimensionalen Raum abgebildet.

nissen auf Basis von 5 verschiedenen Digitalkameramodellen [KSM04] beziehungsweise auf Basis von 13 verschiedenen Handykameramodellen [ÇSA08], lassen sich anhand der Bilddatenbank mit 25 verschiedenen Kameramodellen und insgesamt 73 Digitalkameras deutlich repräsentativere Aussagen ableiten. Hinzu kommt, dass anders als in der bisherigen Literatur Bilder mit unterschiedlichen Kameraeinstellungen aufgenommen und verwendet wurden.

Abbildung 5 zeigt exemplarisch den Abstand von unterschiedlichen Digitalkameras im Merkmalsraum. Für die Erstellung der Abbildung wurden zunächst die zehn wichtigsten Merkmale ausgewählt. Anschließend wurden die Mittelwerte der Merkmalsvektoren einer Kamera durch multidimensionale Skalierung in einen zweidimensionalen Raum abgebildet. In der Abbildung ist zu sehen, dass Kameras des gleichen Modells im Allgemeinen sehr eng beieinander liegen und eine Trennung von Kameras anderer Modelle möglich ist. Dabei gibt es aber auch Ausnahmen. So ist der Abstand zwischen den beiden Geräten des Kameramodells Nikon D200 deutlich größer als bei anderen Kameramodellen.

Zusätzlich sind in Abbildung 6 die ersten beiden Hauptkomponenten für alle Bilder der beiden Kameramodelle Nikon Coolpix S710 und Canon Ixus 70 dargestellt. Während eine Trennung von zwei unterschiedlichen Kameramodellen auf Basis der berechneten Merkmale möglich ist, können Geräte eines Modells auf Basis dieser Charakteristiken nicht unterschieden werden. Die Ergebnisse veranschaulichen erstmalig die erzielbaren Ergebnisse der Kameramodellidentifikation auf Basis des Verfahrens von Kharrazi et al. anhand einer großen Menge von Geräten. Eine ausführlichere Beschreibung der Untersuchungen und der erzielten Ergebnisse auf Basis von Kameramenge A ist in Referenz [GBW09] zu finden.

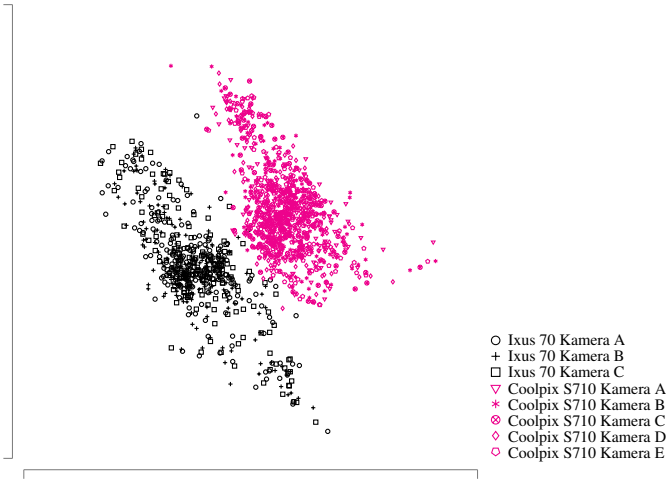

Abbildung 6: Darstellung der Trennbarkeit der beiden Kameramodelle Panasonic DMC-FZ50 (magenta) und Canon Ixus 70 (schwarz) anhand der ersten beiden Hauptkomponenten. Unterschiedliche Geräte sind mit unterschiedlichen Symbolen markiert. Demzufolge ist eine Trennung zwischen Geräten eines Kameramodells auf Basis der berechneten modellabhängigen Merkmale nach der Methode von Kharrazi et al. [KSM04] nicht möglich.

6 Zusammenfassung

Die ‘Dresden Image Database’ wurde als Referenzdatenbank für die Entwicklung und Untersuchung von Methoden der digitalen Bildforensik entwickelt. Die bereitgestellten Bilder decken die typischen Aufnahmesituationen unter Berücksichtigung verschiedener Kameraeinstellungen und verschiedener Umgebungen ab. Die Bilddatenbank ist somit ideal für die Untersuchung von Kameracharakteristiken unter Einbeziehung der wichtigsten Einflussfaktoren geeignet. Generell ermöglicht die Bereitstellung von Forschungsergebnissen auf Basis der Bilddatenbank einen besseren Vergleich von unterschiedlichen Ansätzen und kann das Nachvollziehen, Überprüfen und Beurteilen von Verfahren hinsichtlich ihrer Praxistauglichkeit deutlich vereinfachen. Obwohl die Erstellung der Bilddatenbank speziell unter der Berücksichtigung der Bedürfnisse von Forschern auf dem Gebiet der digitalen Bildforensik vorgenommen wurde, kann sie auch eine nützliche Grundlage für Untersuchungen in verwandten Forschungsgebieten wie beispielsweise der Steganographie bilden.

Die Bilddatenbank wird im Anschluss an die Vorstellung dieses Aufsatzes auf der Informatik 2009 im Internet für Forschungszwecke freigeben. Genauere Informationen werden dann unter folgender Url bekanntgegeben http://dud.inf.tu-dresden.de/~gloe.

Danksagung

Ich bedanke mich bei allen Sponsoren für die zahlreichen Leihgaben von verschiedenen Digitalkameras und insbesondere für die Leihgabe mehrere Geräte eines Kameramodells, namentlich, bei den Mitarbeitern der Fakultät Informatik, den Mitarbeitern des AVMZ (TU Dresden), und folgenden Kameraherstellern (in alphabetischer Reihenfolge): Agfa-Photo & Plawa, Casio, FujiFilm, Kodak, Nikon, Olympus, Pentax, Praktica, Ricoh, Rollei, Samsung und Sony. Weiterhin möchte ich mich für die erteilten Aufnahmegenehmigungen im Zwinger der Stadt Dresden, in den Skulpturensammlungen und in der Orangerie von Schloß Pillnitz bedanken. Besonderer Dank gilt meinen Kollegen, Freunden und Studenten für die vielen Stunden Unterstützung bei der Erstellung der Aufnahmen an einigen der wahrscheinlich kältesten Tagen des Jahres 2009. Eine Realisierung dieses Projekts wäre in dieser Form ohne sie nicht möglich gewesen.

Literatur

[CL01] Chih-Chung Chang und Chih-Jen Lin. *LIBSVM: A Library for Support Vector Machines*. Software verfügbar unter `http://www.csie.ntu.edu.tw/~cjlin/libsvm`, 2001.

[ÇSA08] Oya Çeliktutan, Bülent Sankur und İsmail Avcibas. Blind Identification of Source Cell-phone Model. *IEEE Transactions on Information Forensics and Security*, 3(3):553–566, September 2008.

[GBW09] Thomas Gloe, Karsten Borowka und Antje Winkler. Feature-Based Camera Model Identification Works in Practice – Results of a Comprehensive Evaluation Study. In *11th Information Hiding, Darmstadt, Germany, June 7-10, 2009*, im Druck, 2009.

[GKWB07] Thomas Gloe, Matthias Kirchner, Antje Winkler und Rainer Böhme. Can we Trust Digital Image Forensics? In *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia, September 24–29, 2007, Augsburg, Germany*, Seiten 78–86, New York, NY, USA, 2007. ACM Press.

[HC06] Yu-Feng Hsu und Shih-Fu Chang. Detecting Image Splicing using Geometry Invariants and Camera Characteristics Consistency. In *Proceedings of the 2006 IEEE International Conference on Multimedia and EXPO (ICME 2006)*, Seiten 549–552, 2006.

[JF05] Micah K. Johnson und Hany Farid. Exposing Digital Forgeries by Detecting Inconsistencies in Lighting. In *MM&Sec'05, Proceedings of the Multimedia and Security Workshop 2005, August 1-2, 2005, New York, NY, USA*, Seiten 1–10, 2005.

[JF06] Micah K. Johnson und Hany Farid. Exposing Digital Forgeries through Chromatic Aberration. In *MM&Sec'06, Proceedings of the Multimedia and Security Workshop 2006, September 26-27, 2006, Geneva, Switzerland*, Seiten 48–55, 2006.

[KSM04] Mehdi Kharrazi, Husrev T. Sencar und Nasir Memon. Blind Source Camera Identification. In *Proceedings of the 2004 IEEE International Conference on Image Processing (ICIP 2004)*, Seiten 709–712, 2004.

[LFG05] Jan Lukáš, Jessica Fridrich und Miroslav Goljan. Determining Digital Image Origin Using Sensor Imperfections. In Amir Said und John G. Apostolopoulus, Hrsg., *Proceedings of SPIE: Image and Video Communications and Processing*, Bd. 5685, Seiten 249–260, 2005.

[LFG06a] Jan Lukáš, Jessica Fridrich und Miroslav Goljan. Detecting Digital Image Forgeries Using Sensor Pattern Noise. In Edward J. Delp und Ping Wah Wong, Hrsg., *Proceedings of SPIE: Security and Watermarking of Multimedia Content VIII*, Bd. 6072, 2006.

[LFG06b] Jan Lukáš, Jessica Fridrich und Miroslav Goljan. Digital Camera Identification from Sensor Noise. *IEEE Transactions on Information Forensics and Security*, 1(2):205–214, 2006.

[NCLS06] Tian-Tsong Ng, Shih-Fu Chang, Ching-Yung Lin und Qibin Sun. Passive-blind Image Forensics. In Wenjun Zeng, Heather Yu und Ching-Yung Lin, Hrsg., *Multimedia Security Technologies for Digital Rights*, Kapitel 15, Seiten 383–412. Academic Press, 2006.

[PF04] Alin C. Popescu und Hany Farid. Statistical Tools for Digital Forensics. In Jessica Fridrich, Hrsg., *Information Hiding. 6th International Workshop, IH 2004, Toronto, Canada, May 2004, Revised Selected Papers*, Jgg. LNCS 3200 of *Lecture Notes in Computer Science*, Seiten 128–147, Berlin, Heidelberg, 2004. Springer Verlag.

[SM08] Husrev T. Sencar und Nasir Memon. *Overview of State-of-the-art in Digital Image Forensics*. Statistical Science and Interdisciplinary Research. World Scientific Press, im Druck, 2008.